

Semantic metadata in the publishing industry – technological achievements and economic implications

Tassilo Pellegrini¹

Received: 1 May 2014 / Accepted: 13 October 2016 / Published online: 15 November 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract This paper discusses the adoption of new data management practices, known as Linked Data, by publishing companies from a strategic management perspective. It investigates the role of semantic metadata in the creation and exploitation of so called Linked Data ecosystems and discusses organizational effects on business and value creation processes stimulated by the technology. The theoretical assumptions are complemented by two case studies that provide insights how publishing companies use Linked Data technologies to diversify their business, strategically position themselves within emerging business ecosystems and adapt to the affordances of advanced data management practices. The analysis reveals that despite of differences in the adoption strategy both publishing companies have identified Linked Data technologies as the core of their innovation strategy having a profound impact on existing business practices and new strategies of value creation.

Keywords Linked data · Content value chain · Editorial workflows · Semantic web · Data licensing

JEL Classification L63 · L82 · L86 · O30 · O34

Introduction

Technological progress in the production, distribution and consumption of digital content has nurtured a business environment, in which the cost-efficient handling of internal and external data has become critical for innovations at the process- and product-level (Müchner Kreis 2009; The Economist 2010). This motivates the question if and how the publishing industry makes use of advanced data management technologies to diversify their service and product portfolio, organizationally adapt to the affordances of new data management practices and utilizes them to position itself strategically in the emerging ecosystem of data clouds.

The automation of editorial workflows, i.e. to support dynamic content publishing (Rayfield 2012), and the increasing proliferation of machine-generated content, i.e. as exercised by data journalism (Chambert and Gray 2012), causes and stimulates the adoption of new data management technologies which support the time critical and context-sensitive creation, curation and marketing of digital content (Zwick and Knott 2009). Given the fact that from the vast amount of digital content produced every day just about 5 % is “structured” (Russom 2011), new data management practices are being implemented to improve the machine-processability of digital content. This is achieved not just by creating and enriching content with or on top of structured data, but also by applying metadata standards that support

Responsible Editors: Hans-Dieter Zimmermann and Artur Lugmayr

✉ Tassilo Pellegrini
tassilo.pellegrini@fhstp.ac.at

¹ Department of Media Economics,
University of Applied Sciences St. Pölten,
Matthias Corvinus Strasse 15, A-3100 St. Pölten, Austria

interoperability at a syntactic and semantic level. One of these approaches is called Linked Data.¹

Linked Data is a generic technology to structure and query federated data, thus enabling the flexible and cost efficient reutilization of dispersed digital assets (Cranford 2009). Case studies² from various industries reveal that Linked Data fits well into the incremental IT development practices of enterprises and public organisations, but additionally entails disruptive organisational and institutional effects that pose significant challenges to and opportunities for business diversification (Archer et al. 2013; Pellegrini 2014).

The aim of this paper is twofold. First, it approaches the topic from a strategic management perspective discussing the role of Linked Data technologies in the diversification of product and service portfolios. To do so, the paper provides insights into how Linked Data contributes to the content value chain, identifies typical stakeholder roles within Linked Data ecosystems and discusses asset types and licensing policies in the commercialization of Linked Data assets. Second, the theoretical assumptions laid out in this paper are complemented by two case studies that illustrate the adoption strategy of Linked Data technologies at two publishing companies. The findings help to better understand the organisational impact of new data management practices from a resource-based point of view (Barney 1995) and the role of Linked Data technologies as a service infrastructure for data-driven diversification and the creation of competitive advantage within so called service ecosystems (Frow et al. 2014; Chandler and Lusch 2015). These shall be understood as “relatively self-contained, self-adjusting systems of resource-integrating actors connected by shared institutional logics and mutual value creation through service exchange” (Vargo and Lusch 2011, p. 11).

The paper is structured as follows: Section 2 illustrates the changing role of metadata in the publishing industry and highlights the added value of Linked Data as a new paradigm to turn data into a network good. Section 3 discusses the Linked Data ecosystem by differentiating prototypical data traffic patterns, the contribution of Linked Data to the content value chain and associated licensing issues. Section 4 presents two use cases from the publishing industry to illustrate how these companies have positioned themselves within the Linked Data ecosystem. Section 5 gives a conclusion and outlook on future research.

¹ For a detailed description of the Linked Data approach see also <http://www.w3.org/standards/semanticweb/data>, accessed July 5, 2015

² See, i.e. Rayfield (2012) for a detailed BBC use case or Dodds and Davis (2009) for guardian.co.uk. Additional use cases can be found in Cardoso et al. (2008), Wood (Wood 2010 & Wood 2011), Pellegrini et al. (2014). See also the collection of use cases of the W3C: <http://www.w3.org/2001/sw/sweo/public/UseCases/>, accessed February 22, 2014

Metadata and the publishing industry

Metadata from a resource-based point of view

Industrialization and – as a consequence – the digitization of information production coincided with the increasing proliferation of metadata management as a critical factor in the organization of large quantities of coded information (Bowker and Star 1999; Kitchin 2014) and the diversification of product and service portfolios on top of existing assets (Hass 2011; Pellegrini 2013). From a resource-based perspective (Barney 1995 & Barney 2001; Barney et al. 2011) building, advancing and utilizing metadata capabilities for the diversification of product and service portfolios becomes a strategic resource and a core component to the profit maximization strategies of publishing companies. Building on the argument of Sirmon et al. (2011), the strategic utilization of metadata can be interpreted as a means of resource orchestration, especially when the business environment of the company is characterized by supply inelasticity and the company itself exercises control over the distribution channels for its products and services. Under such circumstances metadata is being applied to reutilize and recompile existing digital assets for the cost-efficient creation and marketing of new products and services, and the deliberate exploitation of market opportunities. Hence, the professionalization of metadata management should be understood as a strategic activity that generates valuable, non-imitable resources which form the basis for new business practices and competitive advantage – either by reducing operational costs or by extending strategic capabilities into new markets. Herein, the author follows the argument of Wan et al. (2011, p. 1340) that it is not “external market failure [that] encourages firms to engage in internal growth; rather, it [is] an internal resource perspective that underscores firms’ motivation to maximize their resources by diversifying into (related) businesses”. The following sections will discuss the increasing importance of and contemporary trends in metadata management and the strategic implications of Linked Data as a technological enabler for resource maximization and business diversification.

Towards a metadata shift

Originating from the library and information sciences metadata standards and practices have spilled over into other industries increasingly being utilized for two purposes: the semantic description and the automated exchange of data. One of the first to adopt progressive metadata practices was the newspaper industry that at the verge of digitization in the 1960s started to develop unified exchange formats for news content

of all common media types.³ About a decade later the book industry started to adopt and develop specific metadata standards for broader purposes of media asset management nowadays known as MARC (Machine Readable Cataloging),⁴ ONIX for Books,⁵ DublinCore⁶ or PRISM (Publishing Requirements for Industry Standard Metadata)⁷ to name, but a few.

With the emergence of the World Wide Web as a universal platform for the creation and distribution of information the nature and functional characteristics of metadata have changed significantly. This trend is illustrated by a survey conducted by Saumure and Shiri (2008) on research topics in the Library and Information Sciences before and after 1993. Table 1 shows their research results.

The survey illustrates three trends: 1) the spectrum of research areas has broadened significantly, 2) while certain areas have kept their status over the years (i.e. Cataloging & Classification or Machine Assisted Knowledge Organization), new areas of research have entered the discipline (i.e. Metadata Applications & Uses, Classifying Web Information, Interoperability Issues) while others have declined or dissolved into other areas (i.e. Cognitive Models), and 3) practical aspects of metadata management have become the primary research area.

The “Metadata Shift” (Haase 2004) described in the survey mentioned above can also be observed in the media industry and related initiatives (i.e. Dublin Core, who started its activities in 1994⁸) that, since the 1990s, have begun to increasingly address issues of interoperability in the exchange of data. The standards IPTC NewsCodes⁹ and the recently introduced semantic web-enabled microformat rNews v1.0¹⁰ are examples of this trend. Both metadata standards are composed of a reasonably manageable amount of concept classes with a sufficient level of semantic expressivity in terms of domain-specific vocabulary and data types that cover the most important attributes of a news item and related media types. The metadata standards can be extended with any other controlled vocabulary (i.e. DublinCore) that adhere to Semantic Web

Table 1 Changes in research areas of the library and information science. Source: Saumure and Shiri (2008)

Research Area	Pre-Web	Post-Web
Metadata Applications / Uses	--	16 %
Cataloguing/ Classification	14 %	15 %
Classifying Web Information	--	14 %
Interoperability	--	13 %
Machine Assisted Knowledge Organization	14 %	12 %
Education	7 %	7 %
Digital Preservation/ Libraries	--	7 %
Thesauri Initiatives	7 %	5 %
Indexing/ Abstracting	29 %	4 %
Organizing Corporate or Business Information	--	4 %
Librarians as Knowledge Organizers of the Web	--	2 %
Cognitive Models	29 %	1 %

standards (like the Resource Description Format - RDF¹¹). By applying Semantic Web principles to existing metadata standards, the media industry is incrementally provided with a technological infrastructure that improves the machine-readability and semantic interoperability of digital media assets, and also changes their good characteristics from isolated artefacts into interconnected assets within a digital ecosystem.

Technological impact of linked (meta) data

Semantic interoperability is crucial to building cost-efficient, interconnected IT systems that integrate numerous data sources (Cranford 2009; Mitchell and Wilson 2012). Since 2009, the Linked Data paradigm has emerged as a light weight approach to improve data portability between various systems on top of a common data model called RDF (Resource Description Framework). By building on RDF (and additional Semantic Web standards like OWL¹² or SPARQL¹³), the Linked Data approach offers significant benefits compared to conventional data management practices. These are according to Auer (2011):

- De-Referencability. Identifiers (URIs) are not just used for identifying entities, but since they can be used in the same way as URLs they also enable locating and retrieving resources describing and representing these entities on the Web.
- Coherence. When an RDF triple contains URIs from different namespaces in subject and object position, this triple establishes a link between the entity identified by the subject (and described in the source dataset using

³ Beside the IPTC news-relevant metadata sources are also provided by the Newspaper Association of America (i.e. ANPA-1312, a 7-bit news agency text markup) or the microformat hNews as provided by Associated Press. Additional standards commonly used in the media industry are EXIF, Dublin Core, XMP, DIG35, P/Meta, ETSI/TV Anytime or MPEG-7. Nevertheless, most of them are designed for multimedia purposes and play a minor role in the news production process.

⁴ See <http://www.loc.gov/marc/bibliographic/>, accessed August 1, 2015

⁵ See <http://www.editeur.org/8/ONIX/>, accessed August 1, 2015

⁶ See <http://dublincore.org/>, accessed August 1, 2015

⁷ See <http://www.idealiance.org/specifications/prism-metadata-initiative>, accessed August 1, 2015

⁸ See <http://dublincore.org/about/history/>, accessed August 1, 2015

⁹ The IPTC NewsCodes consist of the following subsets: EventsML-G2, NewsML-G2, SportsML-G2, rNews, IIM, NewsML 1, IPTC 7901, NITF. See <https://iptc.org/>, accessed August 1, 2015

¹⁰ See: <http://dev.iptc.org/rNews>, visited April 4, 2015

¹¹ See: <http://www.w3.org/RDF/>, visited April 4, 2015

¹² See: <http://www.w3.org/2001/sw/wiki/OWL>, visited April 4, 2015

¹³ See: <http://www.w3.org/TR/rdf-sparql-query/>, visited April 4, 2015

namespace A) with the entity identified by the object (described in the target dataset using namespace B). Through these typed RDF links, data items are effectively and coherently interlinked.

- **Integrability.** Since all Linked Data sources share the RDF data model, which is based on a single mechanism for representing information, it is very easy to attain a syntactic and simple semantic integration of different Linked Data sets. A higher-level semantic integration can be achieved by employing schema and instance matching techniques and expressing found matches again as alignments of RDF vocabularies and ontologies in terms of additional triple facts.
- **Timeliness.** Linked Data can be easily published and updated, and thus facilitates a timely availability. In addition, once a Linked Data source is updated it can instantaneously be accessed and used, since time consuming and error-prone extraction, transformation and loading is not required.

On top of these technological principles Linked Data promises to improve primarily the reusability and richness (in terms of depth and broadness) of digital content, but also altering traditional editorial workflows towards new forms of resource integration. Based upon this approach, the following section will elaborate how Linked Data can be utilized to add value to the content value chain in the production and distribution of digital content.

The linked data ecosystem

Linked Data marks a transition from hierarchies to networks as an organisational principle for information. Hence, the primary value proposition of Linked Data is rooted in its modular flexibility and network characteristics deriving thereof. By sharing the Resource Description Framework (RDF) as a common data model, Linked Data provides the infrastructure for publishing and repurposing of data on top of semantic interoperability.

Linked data traffic patterns

Taking the network characteristics of Linked Data into account, it is possible to identify three prototypical usage scenarios that leverage the potential of increased semantic interoperability, here described as Linked Data Traffic Patterns.

Scenario 1: Internal Perspective: From an internal perspective organizations utilize Linked Data principles to organize information within closed organizational settings. This is especially relevant for organizations whose information is spread among dispersed databases or repositories, entailing challenges with respect to integrating and querying federated data and the harmonisation of legacy issues. Linked Data is bearing a high potential in consolidating dispersed

information infrastructures without necessarily disrupting existing systems and workflows.

Scenario 2: Inbound Perspective: In the second scenario organizations aggregate data from external data sources for purposes like content pooling or content enrichment. This trend is basically backed by the increasing availability of open data, i.e. provided by governmental bodies, community projects (i.e. Wikipedia,¹⁴ Musicbrainz¹⁵ or Geonames¹⁶) or enterprises (i.e. Socrata,¹⁷ Factual¹⁸ or Datamarket¹⁹). Instead of creating these resources on their own, organizations can reuse existing data according to the Terms of Trade of the rights holder – sometimes free of charge or sometimes as paid service according to the service levels of an application programming interface (API).

Scenario 3: Outbound Perspective: In the third scenario organizations apply Linked Data principles to publish data on the web either as open data or via an API that allows the retrieval of data according to predefined service level agreements. This process called Linked Data Publishing is basically a diversification of a data distribution strategy and allows an organization to become part of a Linked Data Cloud (Halford et al. 2012; Jentzsch 2014). Data publishing strategies often go hand in hand with the diversification of business models and require a good understanding of the licensing issues associated with it (Pellegrini 2014).

Linked data in the content value chain

The value chain approach, as introduced by Michael Porter (1985/ Porter 1998) in the mid 1980s, is a core concept in strategic management which describes the structure of sector specific value creation mechanisms and sequential production logics. It has been adopted in a variety of ways for the information industry (i.e. Zerdick 2000; Kim et al. 2004), and recently, it has also gained popularity to systematize the value creation process in data-driven business models as part of the European Commission's Open Data Initiative (COM/2010/0245 final).

The value chain is comprised of distinct stages in the process of value creation, from which each step, in various degrees, contributes to the competitive advantage of an organization. According to this concept the content value chain consists of five steps: 1) content acquisition, 2) content editing, 3) content bundling, 4) content distribution and 5) content consumption. As illustrated in Fig. 1, Linked Data can contribute to each step by supporting the associated intrinsic production function.²⁰

¹⁴ See also www.wikipedia.org, accessed May 20, 2015

¹⁵ See also <http://classic.musicbrainz.org/>, accessed May 20, 2015

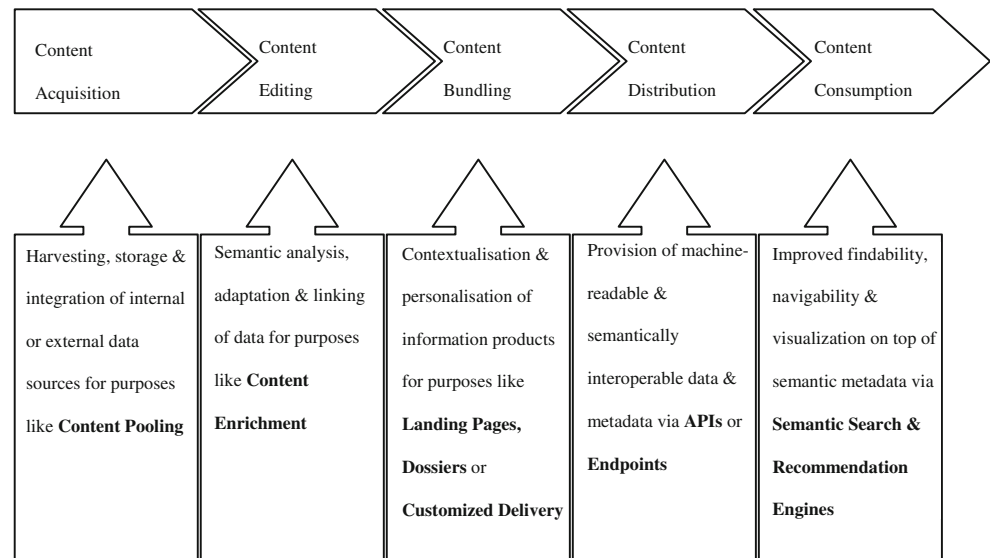
¹⁶ See also <http://www.geonames.org/>, accessed May 20, 2015

¹⁷ See also <http://www.socrata.com/>, accessed May 20, 2015

¹⁸ See also <http://www.factual.com/>, accessed May 20, 2015

¹⁹ See also <https://datamarket.com/>, accessed May 20, 2015

²⁰ A detailed literature review on this topic is provided by Pellegrini (2012).

Fig. 1 Linked data in the content value chain

Content acquisition mainly comprises the collection, storage and integration of relevant information necessary to produce a marketable product or service. In the course of this process, all necessary components are pooled from internal or external sources for further processing. Recent developments have illustrated that Linked Data has been successful in tackling the problem of automated content aggregation (Graube et al. 2011; Hee et al. 2007; Heino et al. 2011), especially in connection with multimedia information (Schandl et al. 2011; Messina et al. 2011), and its enrichment with data from Linked Data sources like DBpedia²¹ or the Linked Movie Data Base.²² Hausenblas (2009), Kobilarov et al. (2009) and Rayfield (2012) provide a comprehensive insight into how the BBC is pulling Linked Data to improve existing web applications for purposes such as content syndication, enrichment and page navigation. I.e. BBC Music is aggregating data from MusicBrainz, Wikipedia and DBpedia to enrich its own database with external information on music related topics. And BBC Sport is using Linked Data technologies to aggregate data from hundreds of internal sources for the automatic generation of landing pages for individual athletes, teams, sports disciplines and competitions.

Content editing includes all necessary steps that deal with the adaptation, interlinking and enrichment of data. Adaptation can be understood as a process, in which acquired data is organized and provided in a way, so that it can be used in the editorial process. Interlinking and enrichment are often performed via processes like tagging and/or referencing of other media assets. Content editing is a highly time- and cost-intensive activity. Hence, cost-efficiency and quality considerations are at the core of the content editing process and

potential areas for automisation on top of expressive metadata. Early work (Kosch et al. 2005; Ohtsuki et al. 2006; Smith and Schirling 2006) provides design principles for a metadata life cycle management and demonstrates the value of well-structured metadata for indexing and compiling multimedia documents across various modalities like text, speech and video. More recent approaches investigate the benefits of ontologies in organising and reutilising semantic metadata for purposes such as metadata enrichment (Hu et al. 2009; Mannens et al. 2009), collaborative tagging practices (Kim et al. 2008) or adaptive content services (Yu et al. 2010).

Content bundling mainly comprises the compilation, contextualisation and personalisation of information products. It can be used to provide customized access to media files, i.e. by using metadata for the device-sensitive delivery of media assets, or to compile thematically relevant material into comprehensive products or product lines, improving thus the navigability, findability and reuse of information. This can be achieved by applying so called mixed-initiative approaches, where machines and humans are interacting in feedback loops when compiling a product or service (i.e. Jokela et al. 2001; Bomhardt 2004; Zhou et al. 2007; Gao et al. 2009; Knauf et al. 2011; Malheiros et al. 2012). Expressive metadata also stimulates purely algorithmic compilation of personalized products (i.e. Liu et al. 2007; Bouras and Tsogkas 2009; Schouten et al. 2010; Ijntema et al. 2010; Goosen et al. 2011) by calculating similarities between media assets, and thus improve the relevance of automated filtering and recommendation services on top of legacy data. This allows new forms of knowledge discovery and delivery services that go beyond the established search and retrieval paradigms and provide the users with a richer interaction experience, often at cost of privacy intrusion and disclosure of personal information.

In a Linked Data environment the process of *content distribution* mainly deals with the provision of machine-

²¹ See: <http://dbpedia.org>, visited April 10, 2015

²² See: <http://www.linkedmdb.org/>, visited April 10, 2015. See also Consens (2008).

readable and semantically interoperable (meta)data via Application Programming Interfaces (APIs) or SPARQL Endpoints (Knowles 2002; Zimmermann 2011). These can be designed either to serve internal purposes, so that data can be reused within the controlled settings of an organization, or for external purposes, so that data can be shared between organizations or with the public. Lots of media-related datasets are already available as Linked Data (i.e. LinkedMDB,²³ DBpedia²⁴ or MusicBrainz²⁵). Over the past years, several media companies have started to offer Linked Data to the public. Since 2009, BBC is offering a SPARQL Endpoints for their program, music, and sports data (i.e. Smethurst 2009; Kobilarov et al. 2009; Rayfield 2012), and in the same year the New York Times has started to offer large amounts of subject headings via its Article Search API²⁶ (Larson and Sandhaus 2009). Similar activities are carried out by Reuters²⁷ and The Guardian²⁸ (Dodds and Davis 2009) or Nature Publishing.²⁹

Content consumption is the last step in the content value chain. This includes any means that enable a human user to search for and interact with media assets in a comfortable and purposeful way. So, according to this view, this level mainly deals with end user applications that make use of Linked Data to provide access to products and services, i.e. by providing reasonable interfaces. Over the past years, increasing attention has been paid to visualization and interaction issues associated with Linked Data although this area of research is still in its infancy (Böhm et al. 2010; Fu et al. 2007; Hoxha et al. 2011; Paulheim 2011; Freitas et al. 2012). Research on and the improvement of interface design for the handling of semantic data services will be one of the critical success factors in the broad adaptation of Linked Data in publishing industry.

Stakeholder roles in a linked data ecosystem

As illustrated in Fig. 2, Latif et al. (2009) propose a simple model that describes the value creation process in a Linked Data ecosystem.

The model distinguishes between various stakeholder roles in the creation of Linked Data assets and various types of data and applications that are created along the data transformation process. Stakeholder roles are raw data provider, linked data provider, application provider and finally end user. The

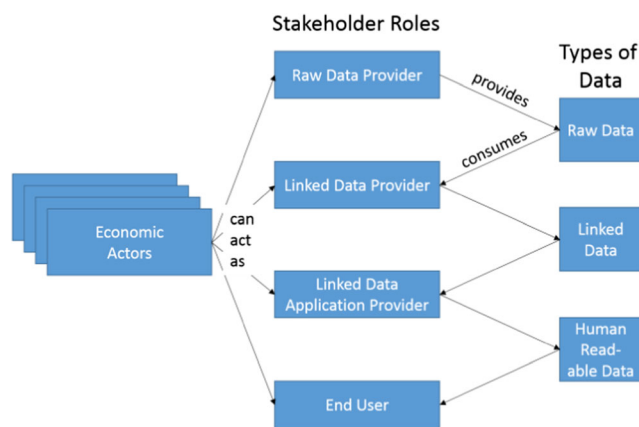


Fig. 2 Linked data value chain. Source: Latif et al. (2009)

stakeholders differ according to their contribution to the Linked Data value chain. Along this process of value creation, raw data – which is provided in any kind of Non-RDF format (i.e. XML, CSV, PDF, HTML etc.) – is transformed into Linked Data which is consumed and processed by a Linked Data application. Finally the end user consumes the human readable data via functionally extended applications and services. As illustrated in Fig. 2, the process of Linked Data creation can be covered in its entirety by a single economic actor or it can be split among several actors who are functionally intertwined via a Linked Data ecosystem. Kinnari (2013) extends this view with an orthogonal layer called “support services and consultation”, stressing the fact that apart from the value creation process itself, Linked Data also creates an environment for added value services that transcends the pure transformation and consumption of data. Such services are usually provided by data brokers who collect, clean, visualize, and resell available data for further processing and consumption.³⁰

For the time being, it is difficult to estimate the cost-effectiveness of Linked Data, but several surveys indicate that depending on the scale and scope of a Linked Data project the saving potential can be significant (Cranford 2009; McHugh 2009). These savings result from the network effect which Linked Data generates as an integration layer across various components and workflows in heterogeneous IT systems. Herein, Linked Data can help to reduce technological redundancies, and thus reducing maintenance costs, improving information access in terms of reduced search and discovery efforts, and providing opportunities for business diversification due to the higher granularity and increased connectivity of digital assets (Mitchell and Wilson 2012).

²³ See <http://www.linkedmdb.org/>, visited April 20, 2015

²⁴ See <http://dbpedia.org/About>, visited April 20, 2015

²⁵ See <http://musicbrainz.org/>, visited April 20, 2015

²⁶ See <http://developer.nytimes.com/docs>, visited April 20, 2015

²⁷ See <http://www.opencalais.com/documentation/calais-web-service-api>, visited April 20, 2015

²⁸ See <http://api.talis.com/stores/guardian>, visited April 20, 2015

²⁹ See <http://www.nature.com/ontologies/>, visited July 15, 2015

³⁰ See also Archer et al. (2013) who carried out a study on business models for Linked Open Government Data. A discussion of the data broker industry is provided by US Committee on Commerce, Science and Transportation (Committee on Commerce, Science and Transportation 2013).

Licensing policies for linked data assets

Technology per se has never been a sufficient precondition for new modes of value creation (Knowles 2002). In the case of Linked Data it is not just the methodology that entails a disruptive potential, but also the changing nature of data as an economic good and its appropriate protection as intellectual property.

Linked Data is comprised of various asset types that emerge during semantic processing of data. These are instance data, metadata, ontologies, services, and technologies. Each asset type contributes in its special way to the value creation process, and thus can be protected by appropriate licensing instruments like copyright, database rights or patents. Table 2 provides an overview over Linked Data assets and related property rights.³¹

The legal regimes of Copyright,³² Database Right,³³ Competition Law³⁴ and Patent Law³⁵ are being complemented by open licensing policies.³⁶ Creative Commons³⁷ allows to define tired licensing policies for the reuse of work protected by copyright. Open Data Commons³⁸ does the same thing for assets protected by database right. And open source licenses complement the patent regime as an alternative form of resource allocation and value generation in the production of software and services (Ghosh et al. 2006).

The open and non-proprietary nature of Linked Data design principles allow to easily share and reuse this data for collaborative purposes. This also offers opportunities to publishers to diversify their assets and nurture new forms of value creation (i.e. by open innovation policies) or unlock new revenue channels (i.e. by establishing highly customizable data syndication services on top of fine granular accounting services). To meet these requirements, commons-based licensing approaches like Creative Commons or Open Data Commons

Table 2 Linked data assets and related property rights

	Copyright	DB Right	Comp. Law	Patents
Instance Data	NO	YES	PARTLY	NO
Metadata	NO	YES	YES	NO
Ontology	YES	YES	YES	NO
Content	YES	NO	YES	NO
Service	YES	NO	YES	PARTLY
Technology	YES	NO	YES	PARTLY

have gained popularity over the last years, allowing reusability while at the same time providing a framework for protection against unfair usage practices and rights infringements. Nevertheless, to meet the requirements of the various asset types, a Linked Data licensing policy should make a deliberate distinction between assets that are protected by database rights and assets that are protected by copyright. Additionally the policy should provide its information not just in a human-readable representation, but also provide the licensing information in a machine-readable way, given the fact that with the increasing reusability of Linked Data the transaction costs of rights clearance and contracting tend to rise. Automated brokering systems can make use of machine-processable licensing information, and thus decreasing the contracting costs significantly. Nevertheless, appropriate licensing practices in the commercial utilization of Linked Data are still in its infancy, but the awareness about the importance of licensing policies as a trigger for business development is rising (Ermilov and Pellegrini 2015).

Linked Data licensing policies provide a secure and resilient judicial framework to protect against the unfair appropriation of (open) datasets and contribute to the strategic aims of an organization to generate competitive advantages, create added value on top of existing assets and diversify its business models.

Case studies in linked data utilization

This chapter discusses two case studies of Linked Data utilization at the publishing companies Wolters Kluwer³⁹ and Reed Elsevier.⁴⁰ Both companies are global players in the market of scientific publishing and consider themselves to be competitors. Additionally, in both cases the supply situation is characterized by a high degree of inelasticity and both companies exercise tight control over the distribution channels of their products and services. Hence, they perfectly fit into the analytic framework of the resource-based theory as laid out in chapter 2. The information given in the two case studies was gathered from company material and interviews

³¹ A detailed discussion of the protection of Linked Data assets as intellectual property can be found in Pellegrini (2014).

³² See also Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society. See also <http://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32001L0029>, accessed April 20, 2015

³³ See also Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases. See also <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31996L0009:EN:HTML>, accessed April 20, 2015

³⁴ See also Consolidated versions of the Treaty on European Union and the Treaty on the Functioning of the European Union - Official Journal C 326, 26/10/2012 P. 0001–0390. See also <http://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:12012E/TXT&from=EN>, accessed April 20, 2015

³⁵ See also <http://www.epo.org/law-practice/legal-texts/html/epc/1973/e/ar52.html>, accessed April 20, 2015

³⁶ A detailed discussion of licensing issues related to Linked Open Data is provided by Pellegrini (2014) and Ermilov and Pellegrini (2015).

³⁷ See also <http://creativecommons.org/>, accessed May 21, 2015

³⁸ See also <http://opendatacommons.org/>, accessed May 21, 2015

³⁹ See also <http://www.wolterskluwer.com/Pages/Home.aspx>, accessed August 10, 2015

⁴⁰ See also <https://www.elsevier.com/>, accessed August 10, 2015

conducted with company representatives who are in charge of innovation management and the companies' transition to Linked Data technologies. The aim was to gain understanding of the adoption of Linked Data technologies and how they contribute to value creation within the company. Both company representatives were interviewed using the same semi-structured interview scheme, with questions focussing on the motivations for Linked Data deployment, the application area, the contribution to the value chain, and the licensing policy associated with newly emerging data assets. After the interviews have been transcribed and analysed according to the criteria mentioned above, the interviewees received a copy of the analysis for review, clarification, suggestions and updates. Table 3 gives an overview over the central findings.

Case 1: Wolters kluwer

Motivation

The primary motivation for Wolters Kluwer to engage in Linked Data is to reduce costs of the editorial process. This is achieved by reutilizing existing assets either from within the company or from trusted third party sources, reducing the efforts to generate and maintain assets themselves or by sharing existing assets across organizational units. Additionally, Linked Data allows Wolters Kluwer to build functionalities and provide services that have not been possible before, triggering innovation on top of collaborative practices. This stimulates new content exploitation practices by utilizing application programming interfaces and other forms of service-based principles on top of automated content processing.

Application area

Wolters Kluwer uses Linked Data technology to support several editorial processes in the compilation and provision of legal information to their customers. At the time of writing, this mainly takes place within their proprietary syndication platform Jurion©, a professional service that supports legal professionals like lawyers, attorneys, judges or notaries in their daily work. The Jurion© platform provides information

management functionalities like search services, libraries and knowledge management services for the professional handling and processing of legal information.

Value chain

Wolters Kluwer utilizes Linked Data technology along its entire content value chain. To achieve this, they have been engaging in a major change management process that has been successively rolled out among the whole corporation. The semantic metadata is steering the whole production process from acquisition to distribution of content. Additionally, semantic metadata is being applied to improve the customer experience at the content consumption level.

Traffic patterns

At the time of writing, Wolters Kluwer uses Linked Data technologies mainly to aggregate and organize internal information from sources from within the company. Inbound aggregation of content from external sources or the provision of content via outbound practices is being considered to be part of future strategies, but for the time being, this is not a primary strategic aim. Nevertheless, Wolters Kluwer is aware of the strategic value generated especially by outbound practices, i.e. by providing data assets to the public and thus initiating positive feedback loops for the content providers in the sense of open innovation.

Stakeholder role

Wolters Kluwer currently acts as a Linked Data provider, i.e. by providing several Linked Data vocabularies to the public, and as a Linked Data application provider, i.e. by utilizing Linked Data within their Jurion© platform. The strategic focus lies primarily on the second aspect as Wolters Kluwer is keen on exercising a sufficient amount of control over their assets and how they are being utilized by third parties. Providing Linked Data assets to the public is part of this strategy. For the time being inbound activities are of minor relevance given the fact that open data often lacks sufficient licensing and provenance information, thus posing a risk to quality assurance and legal security.

Table 3 Comparison of linked data strategies between Wolters Kluwer and reed elsevier

	Wolters Kluwer	Reed Elsevier
Motivation	Reduction of editorial costs	Product diversification
Application Area	Search & retrieval	Customization
Value Chain Support	All stages	Editing & Bundling
Traffic Patterns	Internal use / Outbound	Internal Use
Stakeholder Role	Linked Data provider / Linked Data application provider	Linked Data provider
Licensing Strategy	Open system approach: Dual Licensing	Closed system approach: Proprietary licensing

Licensing strategy

For the first time in the company history, Wolters Kluwer perceives the diversification of its licensing strategy as an issue. In former times, intellectual property issues were comparably simple. In the future, asset management based on a diversified licensing strategy will be a critical issue in the exploitation of new content markets. At the time of writing, Wolters Kluwer is experimenting with the combination of copyright and Creative Commons licenses to develop a nuanced asset management strategy for the future.

Case 2: Reed Elsevier

Motivation

Reed Elsevier's business rationale is to expand their market share among higher education customers by offering advanced learning and certification services. To do so, they have started to reorganize their editorial workflows by introducing a semantic metadata layer that interlinks the assets managed by formerly separated business units. This shall improve the effectiveness in the creation of new products and services and create opportunities for business development under special consideration of new content delivery platforms and consumption habits.

Application area

Reed Elsevier utilizes Linked Data technologies within their product line Elsevier Optimized Learning Suite® (EOLS). EOLS lets its users create so called learning journeys by allowing teachers and students to arrange learning objects in a consistent and responsive way under special consideration of mobile consumption. Linked Data principles are being applied to automatically extract knowledge artefacts from existing sources, arrange them in didactically meaningful ways and offer them as consistent products to their customers.

Value chain

The current application focus lies on editing and bundling of learning objects for purposes like context-sensitive grouping and compiling of content objects. Semantic relationships are being used to create tight cores of semantically interlinked objects, which are being provided as a consistent product. Additionally, semantic metadata is used for recommendation purposes of semantically related content objects. So semantics is instrumental to creating the product itself, define product families or product lines and support the reutilization of existing assets for various consumption purposes and

responsive formats as an extension to the traditional product lines (i.e. handbooks).

Traffic patterns

For the time being, Reed Elsevier applies Linked Data in a strictly controlled business environment. They use Linked Data technologies internally to support editorial offices in collaboratively creating semantic relationships, semantically enrich datasets and support the quality approval process. Inbound content is retrieved from certified partners only who have to adhere to strict bylaws, policies and quality standards. Currently, no open data is utilized, but Reed Elsevier leaves the option open for the future. Nevertheless, Reed Elsevier retrieves a lot of content from certified partners which makes the inbound process a critical phase in the value creation process. Inbound efforts are almost equal to efforts for internal processes. From the outbound perspective, no data publishing or publishing of vocabularies or ontologies currently occurs. All outbound activities are subject to strict licensing agreements with certified partners where the costs of maintaining shared resources are shared equally among all partners involved. Currently, there are no plans to open up any assets for public use apart from some low level assets that have no strategic value to the company.

Stakeholder role

Reed Elsevier describes itself as a relatively closed system that does not serve an ecosystem in a broader sense. If Linked Data is exposed to the outside world, than it is done in a very controlled fashion. There are no endpoints or APIs available to the public. Any utilization of Linked Data takes place within a strongly controlled B2B environment, but Reed Elsevier is aware that there are opportunities ahead. The strategic aim is to become a data integration partner and service provider within knowledge and research institutions. This entails the creation of a shared knowledge backend that does not only provide Elsevier's products, but also provides the data services to cross-fertilize research and the record of science as a whole by integrating datasets from previously separated sources and sites.

Licensing strategy

Licensing has become a big issue. Traditional contracting models have become dysfunctional for repurposing and disseminating content under circumstances of multi-channel publishing. Reed Elsevier currently faces a severe backlog in the adjustment of existing licenses. All new contracts take new options of repurposing and appropriate compensation models into account, but existing legacies impose certain obstacles to jumpstart new products, introduce new functionalities and

service-levels. Thus, in the future dynamic licensing will be vital to serve various business models, i.e. allowing new charging models. Dual licensing is currently being applied under their open access policies for certain journals, but data licensing is currently not an issue.

Interpretation of results

The two use cases show similarities and differences in the commercial utilization of Linked Data principles. Similarities exist in the utilization of Linked Data technologies for purposes of collaborative data management and reutilizing assets across business units and organizational boundaries. This is a clear indicator that both companies are applying Linked Data technologies to leverage internal resources for the expansion into new markets (Reed Elsevier) or the improvement of existing costs structures in the production and distribution of existing assets (Wolters Kluwer). In both cases Linked Data serves as a technological and an organizational integration layer, impacting existing workflows and working practices, and triggering new products and business models. Nevertheless, differences exist in the exploitation of these opportunities and in the notion and design of the appropriate ecosystem.

Wolters Kluwer is aiming at nurturing an open business environment inspired by the principles of open innovation. They have started to experiment with providing certain resources under a dual licensing policy, thus stimulating community dynamics and collaborative business practices. Feeding into and retrieving input from an open business environment will become more important in the future and a strategic cornerstone of their business development practices. On the contrary, Reed Elsevier relies on strict control mechanisms in governing the value creation process. They utilize a very strict licensing model, arguing that for reasons of quality assurance they need to exercise tight control over their business environment and associated collaborators. Nevertheless, they are aware of the business opportunities offered by open innovation, but have not yet embraced this culture.

Despite the differences, both publishing companies have identified Linked Data technologies as the core of their innovation strategy having a profound impact on existing business practices and strategies of value creation. It allowed them to use existing resources more effectively and extend their practical capabilities by reutilizing existing resources in new contexts within and beyond the boundaries of their company.

Conclusion & outlook

Without any claim to completeness and representativeness, this paper discusses the strategic utilization of semantic metadata and its impact on business practices in the publishing

industry from a resource-based point of view. Linked Data can be described as a change agent in the strategic transformation of the publishing industry under conditions of digitization and collaborative value creation within service systems. Hence, the resource-based approach discussed at the beginning of this paper provides a robust theoretical framework for the explanation and understanding of the adoption of Linked Data technologies. Nevertheless, numerous issues require further investigation.

One issue is the relative complexity of Linked Data as a technologically induced innovation. The two case studies presented in this paper might claim a blueprint for large enterprises that exercise control over markets with high supply inelasticity, but they have little explanatory value when it comes to the adoption of Linked Data technologies by small and medium sized enterprises. Lack of skills, competencies, financial resources and economies of scale in the production and distribution of content might hinder smaller companies in actively participating in newly emerging service ecosystems. This raises questions about market disparities, anti-competitive behavior, and concentration of market power.

Another aspect that has not been touched upon in this paper is the quality assurance of Linked Data with respect to validity, trustworthiness, and provenance of (open) data. Given the fact that the publishing industry is highly sensitive to such quality criteria, efforts should be undertaken to significantly improve the quality of Linked Data and provide transparent and reliable measures for quality assurance and maintenance – especially when it is created via crowdsourcing or similar collaborative practices. Additionally, it needs to be discussed how free-rider effects can be prohibited and incentives created, not just to consume, but also to contribute to emerging Linked Data ecosystems. Tackling these challenges is equally important as managing the technological feasibility of Linked Data in editorial processes, but probably much harder to accomplish.

A third aspect affects the issue of business and revenue models that appropriately reward the various stakeholders within a Linked Data ecosystem. Linked Data will make a big leap forward if it can prove to significantly reduce the costs of existing editorial workflows, create new revenue opportunities and provide incentives and fair compensation models across the Linked Data value chain. For the time being, these issues are open to debate, although the expectations are very high.

Acknowledgements Open access funding provided by St. Pölten University of Applied Sciences. A preliminary version of this paper (Pellegrini 2012) was presented at the Academic MindTrek Conference 2012, Tampere, Finland, October 3–5, 2012.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Archer, P., Dekkers, M., Goedertier, S., & Loutas, N. (2013). Study on business models for Linked Open Government Data (BM4LOGD - SC6DI06692). Prepared for the ISA programme by PwC EU Services See also: http://ec.europa.eu/isa/documents/study-on-business-models-open-government_en.pdf. Accessed 10 May 2014.
- Auer, S. (2011). Creating knowledge out of interlinked data. In Proceedings of WIMS'11, p. 1–8.
- Barney, J. (1995). Looking inside for competitive advantage. *Academy of Management Executive*, 9(1), 49–61.
- Barney, J. (2001). Resource-based theories of competitive advantage: a ten-year retrospective on the resource-based view. *Journal of Management*, 27(2001), 643–650.
- Barney, J. B., Ketchen, D. J., & Wright, M. (2011). The future of resource-based theory: revitalization or decline? *Journal of Management*, 37(5), 1299–1315.
- Böhm, C., Naumann, F., & Freitag, M. (2010). Linking open government data: What journalists wish they had known. In Proceedings of the 6th International Conference on Semantic Systems, ACM, p. 1–4.
- Bomhardt, C. (2004). NewsRec, a SVM-driven personal recommendation system for news websites. In IEEE/WIC/ACM international conference on web intelligence, pp. 545–548.
- Bouras, C., & Tsogkas, V. (2009). Personalization mechanism for delivering news articles on the user's desktop. In Fourth International Conference on Internet and Web Applications and Services, p. 157–162.
- Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences*. London: MIT Press.
- Cardoso, J., Hepp, M., & Lytras, M. (2008). *The semantic web: Real-world applications from industry (semantic web and beyond)*. New York: Springer.
- Chambert, L., & Gray, J. (2012). *The data journalism handbook*. New York: O'Reilly Media.
- Chandler, J. D., & Lusch, R. F. (2015). Service systems: a broadened framework and research agenda on value propositions, engagement, and service experience. *Journal of Service Research*, 18(1), 6–22.
- Committee on Commerce, Science and Transportation (2013). A review of the data broker industry: Collection, use and sale of consumer data for marketing purposes. Staff Report for Chairman Rockefeller. See also: https://law.ku.edu/sites/law.ku.edu/files/docs/media-law/2014/Data_Broker_Industry_Senate_Report.PDF. Accessed 5 May 2014.
- Consens, M. P. (2008). Managing linked data on the web: The LinkedMDB showcase. In Proceedings of Latin American Web Conference 2008, p. 1–2.
- Cranford, S. (2009). Spinning a data web. In P. W. Coopers (Ed.), Technology forecast. See also: <http://www.pwc.com/us/en/technology-forecast/spring2009/index.jhtml>. Accessed 20 Sept 2012.
- Dodds, L., & Davis, I. (2009). MP Data SPARQL Editor. <http://www.guardian.co.uk/open-platform/apps-mp-data-sparql-editor>. Accessed 20 Apr 2013.
- Ermilov, I., & Pellegrini, T. (2015). Data licensing on the cloud - Empirical evidence and implications. In Proceedings of Semantics 2015 Conference. New York: ACM, p. 153–156.
- Freitas, A., Curry, E., Oliveira, J. G., & O'Riain, S. (2012). Querying heterogeneous datasets on the linked data web. Challenges, approaches, and trends. In IEEE Internet Computing, p. 24–33.
- Frow, P., McColl-Kennedy, J. R., Hilton, T., Davidson, A., Payne, A., & Brozovic, D. (2014). Value propositions: a service ecosystems perspective. *Marketing Theory*, 14(3), 327–351. doi:10.1177/1470593114534346.
- Fu, T.-c., Sze, D. C. M., Leung, P. K. C., Hung, K.-y., & Chung, F.-I. (2007). Analysis and visualization of time series data from consumer-generated media and news archives. In Proceedings of Web Intelligence and Intelligent Agent Technology Workshops (WI-IAT), 2007 IEEE/WIC/ACM, p. 259–262.
- Gao, Feng; Yuhong Li; Li Han; Jian Ma (2009). InfoSlim: An ontology-content based personalized mobile news recommendation system. In 5th International Conference on Wireless Communications, Networking and Mobile Computing, p.1–4.
- Ghosh, R. A. et al. (2006). Economic impact of open source software on innovation and the competitiveness of the Information and Communication Technologies (ICT) sector in the EU. Final report prepared on November 20, 2006 (Contract ENTR/04/112.). See also http://ec.europa.eu/enterprise/sectors/ict/files/2006-11-20-flossimpact_en.pdf. Accessed May 10, 2014.
- Goosen, F., Ijntema, W., Frasinicar, F., Hogenboom, F., Kaymak, U. (2011). News personalization using the CF-IDF semantic recommender. In Proceedings of WIMS'11, p. 1–12.
- Graube, Markus; Pfeffer, Johannes; Ziegler, Jens; Urbas, Leon (2011). Linked data as integrating technology for industrial data. In 2011 International Conference on Network-Based Information Systems, p. 162–167.
- Haase, K. (2004). Context for semantic metadata. In MM'04. New York: ACM.
- Halford, S., Pope, C., & Weal, M. (2012). Digital futures? Sociological challenges and opportunities in the emergent semantic web. *Sociology*, 47/1.
- Hass, B. H. (2011). Intrapreneurship and corporate venturing in the media business: a theoretical framework and examples from the German publishing industry. *Journal of Media Business Studies*, 8(1), 47–68.
- Hausenblas, M. (2009). Exploiting linked data to build web applications. *IEEE Internet Computing*, 13/4, 68–73.
- Hee, K. L., Kim, H. Y., & Lee, H.-K. (2007). News package service based on TV-Anytime metadata gathered from RSS. In IEEE International Symposium on Consumer Electronics, p. 1–6.
- Heino, N., Tramp, S., & Auer, S. (2011). ManagingWeb Content using Linked Data Principles – Combining semantic structure with dynamic content syndication. In 35th IEEE Annual Computer Software and Applications Conference, p. 245–250.
- Hoxha, J., Brahaj, A., Vrandecic, D. (2011). Open.data.al - Increasing the utilization of government data in Albania. In Proceedings of the 7th International Conference on Semantic Systems, ACM, p. 237–240.
- Hu, B., Wang, J., & Zhou, Y. (2009). Ontology design for online news analysis. In WRI Global Congress on Intelligent Systems, p. 202–206.
- Ijntema, W., Goossen, F., Frasinicar, F., & Hogenboom, F. (2010). Ontology-based news recommendation. In EDBT '10 Proceedings of the 2010 EDBT/ICDT Workshops, p. 1–6.
- Jentzsch, A. (2014). Linked open data cloud. In T. Pellegrini, H. Sack, & S. Auer (Eds.), *Linked enterprise data* (pp. 209–220). Berlin: Springer Verlag.
- Jokela, S., Turpeinen, M., Kurki, T., Savia, E., & Sulonen, R. (2001). The role of structured content in a personalized news service. In System sciences. Proceedings of the 34th Annual Hawaii International Conference. p. 1–10.
- Kim, E., Nam, D., & Stimpert, J. L. (2004). The applicability of Porter's generic strategies in the digital age: assumptions, conjectures, and suggestions. *Journal of Management*, 30, 569–589. doi:10.1016/j.jm.2003.12.001.
- Kim, H. L., Passant, A., Breslin, J. G., Scerri, S., Decker, S. (2008). Review and alignment of tag ontologies for semantically-linked data in collaborative tagging spaces. In: IEEE International Conference on Semantic Computing, p. 315–322.
- Kinnari, T. (2013). Open data business models for media industry. A Finnish case study. Master Thesis, Department of Information and Service Economy, Aalto University. See also: http://pub.lib.aalto.fi/fi/thesis/pdf/13166/hse_thesis_13166.pdf. Accessed 29 Mar 2014.

- Kitchin, R. (2014). *The data revolution. Bog data, open data, data infra-structures and their consequences*. London: Sage Publications.
- Knauf, R., Kürsten, J., Kurze, A., Ritter, M., Berger, A., Heinich, S., & Eibl, M. (2011). Produce. Annotate. Archive. Repurpose —: Accelerating the composition and metadata accumulation of content. In Proceedings of the 2011 ACM international workshop on automated media analysis and production for novel TV services, p. 31–36.
- Knowles, C. (2002). Intelligent agents without the hype: why they work best with structured content. *Business Information Review*, 19/4, 22–28.
- Kobilarov, G., Scott, T., Raimond, Y., Oliver, S., Sizemore, C., Smethurst, M., Bizer, C., & Lee, R. (2009). Media meets semantic web – How the BBC uses DBpedia and linked data to make connections. In Proceedings of ESWC 2009, the 6th European Semantic Web Conference (p. 723–747). New York: Springer LNCS.
- Kosch, H., Boszormenyi, L., Doller, M., Libsie, M., Schojer, P., & Kofler, A. (2005). The life cycle of multimedia metadata. *Multimedia, IEEE*, 12(1), 80–86.
- Kreis, M. (2009). Prospects and opportunities of information and communication technologies (ICT) and media. International Delphi Study 2030. Stuttgart: National IT Summit 2009.
- Larson, R., & Sandhaus, E. (2009). NYT to release thesaurus and enter linked data cloud. In <http://open.blogs.nytimes.com/2009/06/26/nyt-to-release-thesaurus-and-enter-linked-data-cloud/>. Visited 20 April 2012.
- Latif, A., Saeed, U. A., Höfler, P., Stocker, A., & Wagner, C. (2009). The linked data value chain: A lightweight model for business engineers. In Proceedings of I-Semantics 2009, the 5th International Conference on Semantic Systems. Graz: Journal of Universal Computer Science, p. 568–577.
- Liu, Y., Wang, Q.-X., Guo, L., Yao, Q., Lv, N., & Wang, Q. (2007). The optimization in news search engine using formal concept analysis. In 4th International Conference on Fuzzy Systems and Knowledge Discovery, p. 45–49.
- Malheiros, M., Jennett, C., Patel, S., Brostoff, S., & Sasse, M. A. (2012). Too close for comfort: A study of the effectiveness and adaptability of rich-media personalized advertising. In Proceedings of CHI 2012, p. 579–588.
- Mannens, E., Troncy, R.; Braeckman, K.; Van Deursen, D.; Van Lancker, W.; De Sutter, R.; Van de Walle, R. (2009). Automatic metadata enrichment in news production. In 10th Workshop on Image Analysis for Multimedia Interactive Services, p.61–64.
- McHugh, L. (2009). Measuring the value of metadata. White Paper: Baseline Consulting.
- Messina, A., Montagnuolo, M., Di Massa, R., & Elia, A. (2011). The hyper media news system for multimodal and personalised fruition of informative content. In Proceedings of ICMR '11.
- Mitchell, I., & Wilson, M. (2012). Linked data. Connecting and exploiting big data. Fujitsu White Paper. <http://www.fujitsu.com/uk/Images/Linked-data-connecting-and-exploiting-big-data-%28v1.0%29.pdf>. Accessed 12 Sept 2012.
- Ohtsuki, K., Bessho, K., Matsuo, Y., Matsunaga, S., & Hayashi, Y. (2006). Automatic multimedia indexing: combining audio, speech, and visual information to index broadcast news. *IEEE Signal Processing Magazine*, 23(2), 69–78.
- Paulheim, H. (2011). Improving the usability of integrated applications by using interactive visualizations of linked data. In Proceedings of the international conference on web intelligence, mining and semantics WIMS '11, ACM (pp. 1–12).
- Pellegrini, T. (2012). Integrating linked data into the content value chain: A review of news-related standards, methodologies and licensing requirements (p. 94). ACM Press. doi:10.1145/2362499.2362513.
- Pellegrini, T. (2013). The economics of big data – A value perspective on state of the art and future trends. In Big data computing (S. 343–372). New York: Chapman and Hall/CRC.
- Pellegrini, T. (2014). Linked data licensing – Datenlizenzierung unter netzökonomischen Bedingungen. In E. Schweighofer, Kummer, F., & Hötendorfer, W. (Hg.). Transparenz. Tagungsband des 17. Internationalen Rechtsinformatik Symposium IRIS 2014. Wien: Verlag der Österreichischen Computerergesellschaft, S. 159–168.
- Pellegrini, T., Sack, H., & Auer, S. (2014). *Linked enterprise data*. Berlin: Springer.
- Porter, M. E. (1998). *Competitive advantage: Creating and sustaining superior performance: With a new introduction, 1st free press*. New York: Free Press.
- Rayfield, J. (2012). Sports refresh: Dynamic semantic publishing. http://www.bbc.co.uk/blogs/bbcinternet/2012/04/sports_dynamic_semantic.html. Visited 20 Apr 2012.
- Russom, P. (2011). Big data analytics. TDWI Research Report, 4. Quarter 2011. <http://tdwi.org/bdr-rpt.aspx>. Accessed 14 Sept 2012.
- Saumure, K., & Shiri, A. (2008). Knowledge organization trends in library and information studies: a preliminary comparison of pre- and post-web eras. *Journal of Information Science*, 34/5, 651–666.
- Schandl, B., Haslhofer, B., Bürger, T., Langedger, A., & Halb, W. (2011). Linked data and multimedia: the state of affairs. *Multimedia Tools and Applications*, 1–34.
- Schouten, K., Ruijgrok, P., Borsje, J., Frasinca, F., Levering, L., & Hogenboom, F. (2010). A semantic web-based approach for personalizing news. In SAC '10 Proceedings of the 2010 ACM Symposium on Applied Computing, p. 854–861.
- Simon, D. G., Hitt, M. A., Ireland, R. D., & Gilbert, B. A. (2011). Resource orchestration to create competitive advantage: breadth, depth, and life cycle effects. *Journal of Management*, 37(5), 1390–1412.
- Smethurst, M. (2009). BBC Backstage SPARQL endpoint for programmes and music. In: http://www.bbc.co.uk/blogs/radiolabs/2009/06/bbc_backstage_sparql_endpoint.html. Visited 20 Apr 2012.
- Smith, J. R., & Schirling, P. (2006). Metadata standards roundup. *IEEE Multimedia*, 13/2, 84–88.
- The Economist (2010). Data, data everywhere. A special report on managing information. <http://www.emc.com/collateral/analyst-reports/ar-the-economist-data-data-everywhere.pdf>. Accessed 30 Mar 2013.
- Vargo, S., & Lusch, R. (2011). It's all B2B and beyond: toward a systems perspective of the market. *Industrial Marketing Management*, 40/2, 181–187.
- Wan, W. P., Hoskisson, R. E., Short, J. C., & Yiu, D. W. (2011). Resource-based theory and corporate diversification: accomplishments and opportunities. *Journal of Management*, 37/5, 1335–1368.
- Wood, D. (Ed.) (2010). *Linking enterprise data*. New York: Springer.
- Wood, D. (Ed.) (2011). *Linking government data*. New York: Springer.
- Yu, H. Q., Benn, N., Dietze, S., Pedrinaci, C., Liu, D., Domingue, J., & Siebes, R. (2010). Two-staged approach for semantically annotating and brokering TV-related services. In IEEE International Conference on Web Services, 497–503.
- Zerdick, A. (2000). European Communication Council (Eds.). Economics: strategies for the digital marketplace: European Communication Council report. Berlin: Springer.
- Zhou, Y.-q., Hu, Y.-f., & He, H.-c. (2007). Learning user profile in the personalization news service. In International Conference on Natural Language Processing and Knowledge Engineering, p. 485–490.
- Zimmermann, A. (2011). Leveraging the linked data principles for electronic communications. In IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, p. 385–388.
- Zwick, D., & Knott, J. D. (2009). Manufacturing customers. The database as a new means of production. *Journal of Consumer Culture*, 9/2, 221–247.